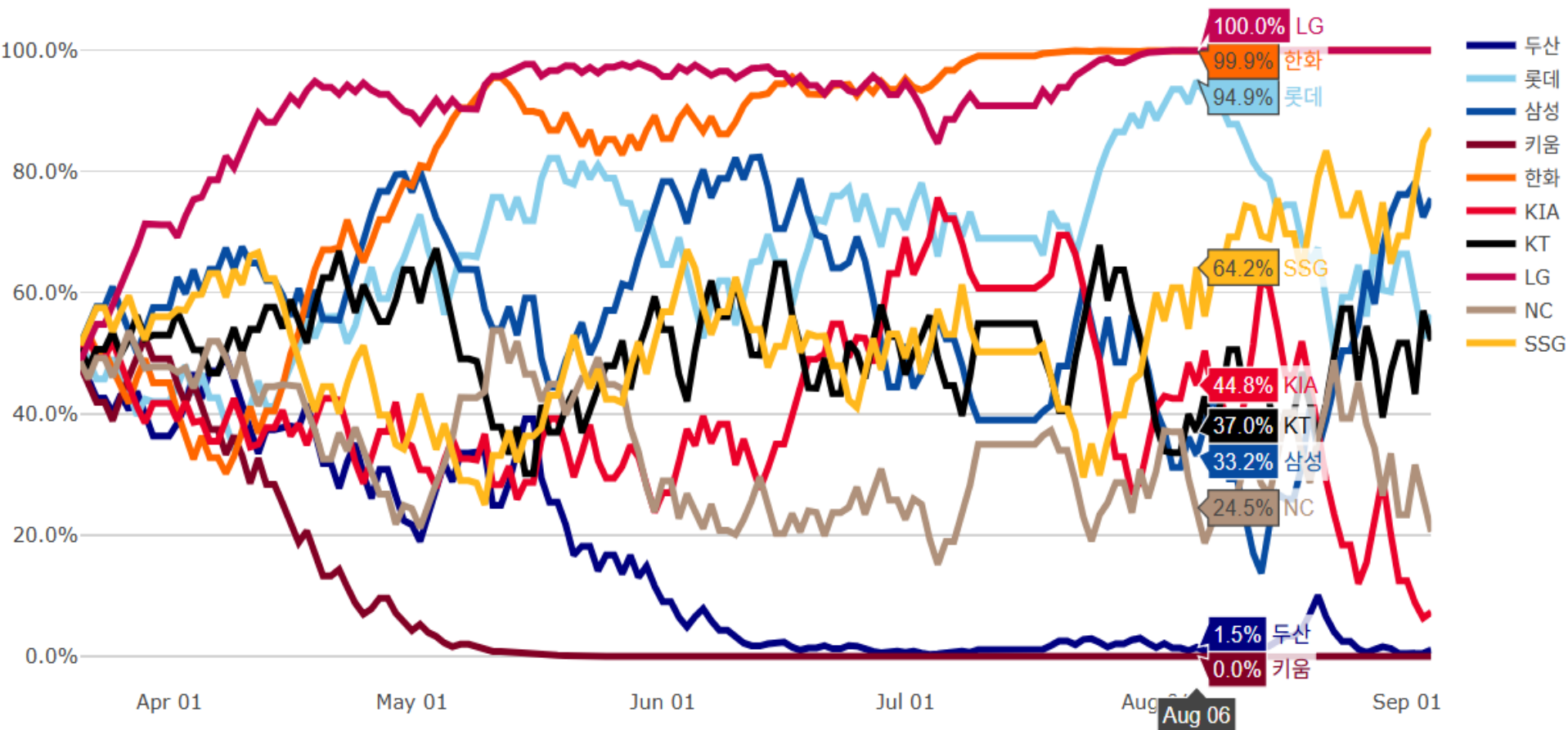


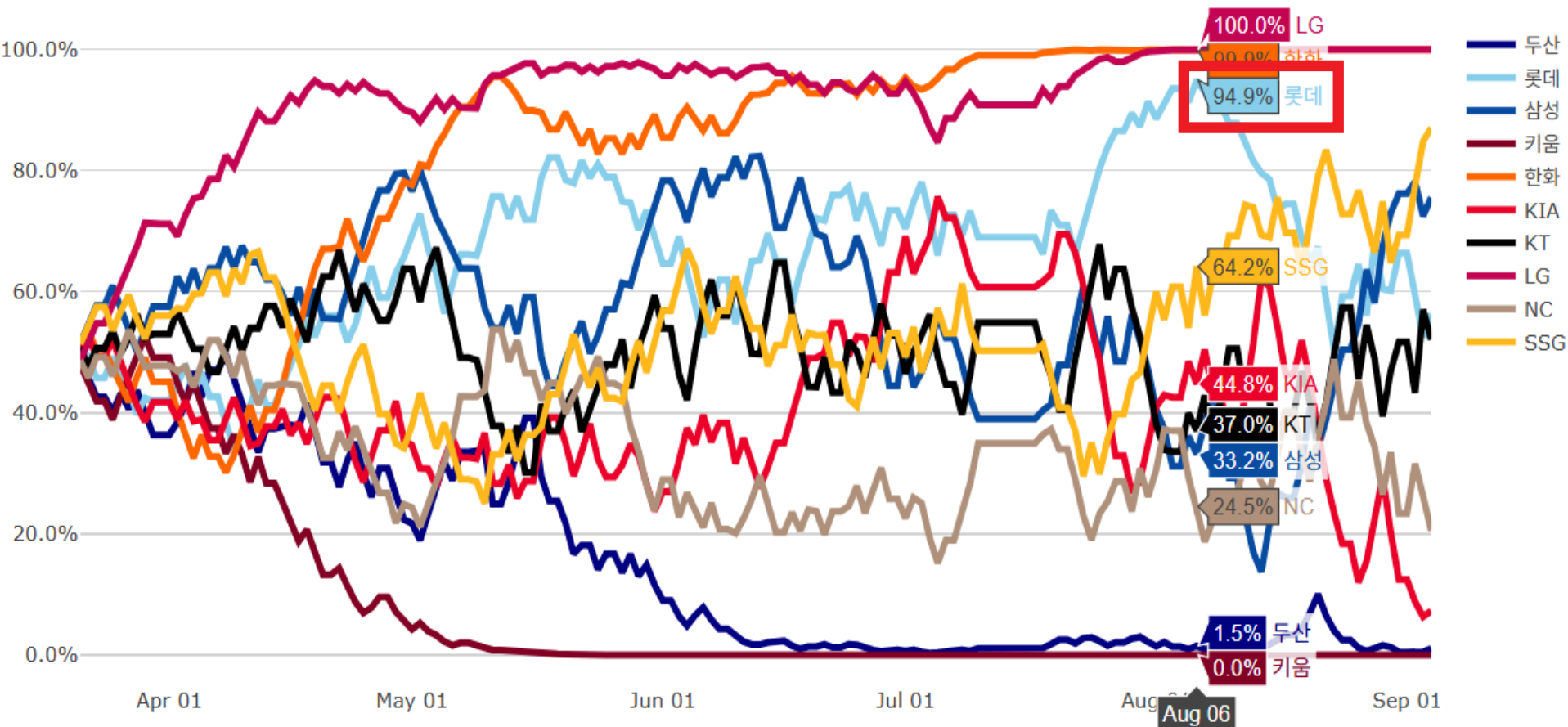
12연패,
누가 범인일까?



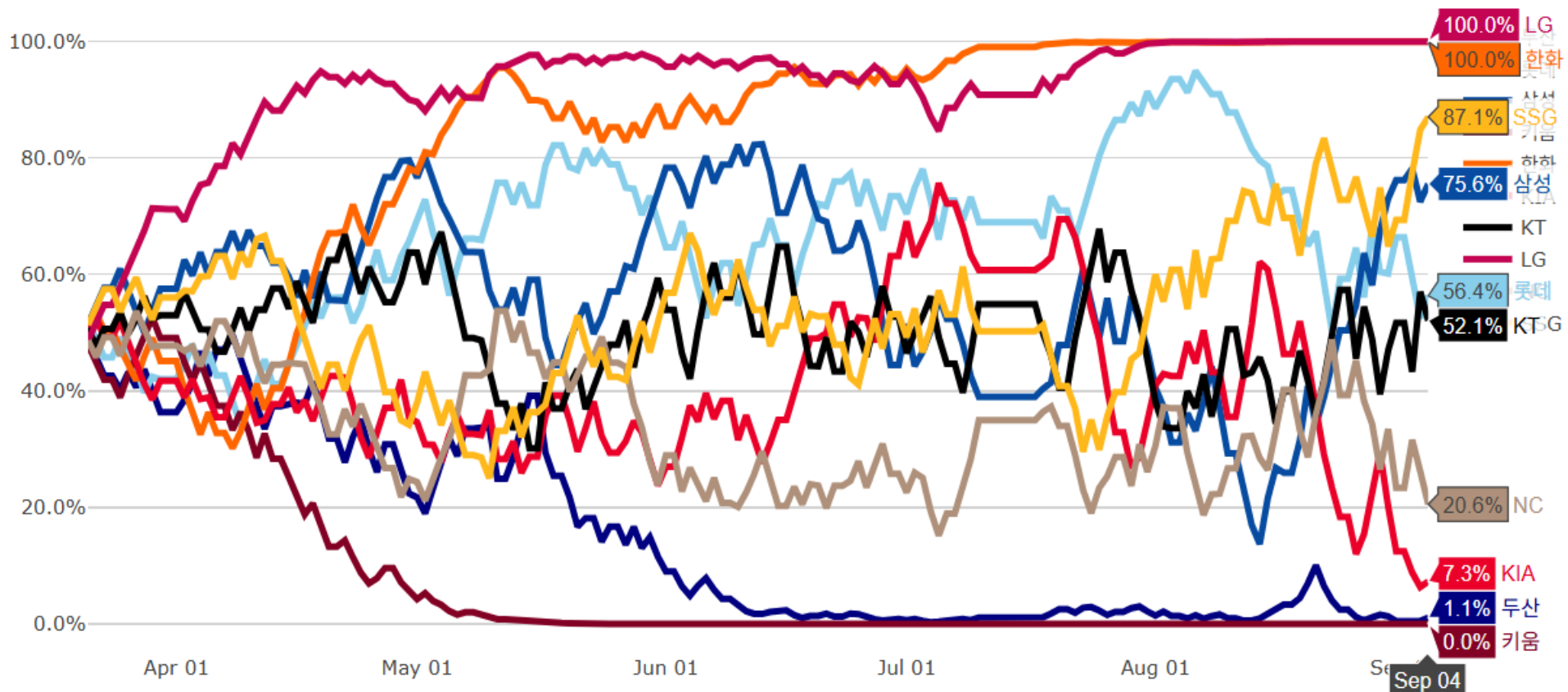
08/06



08/06



09/04



포스트 시즌 진출 확률

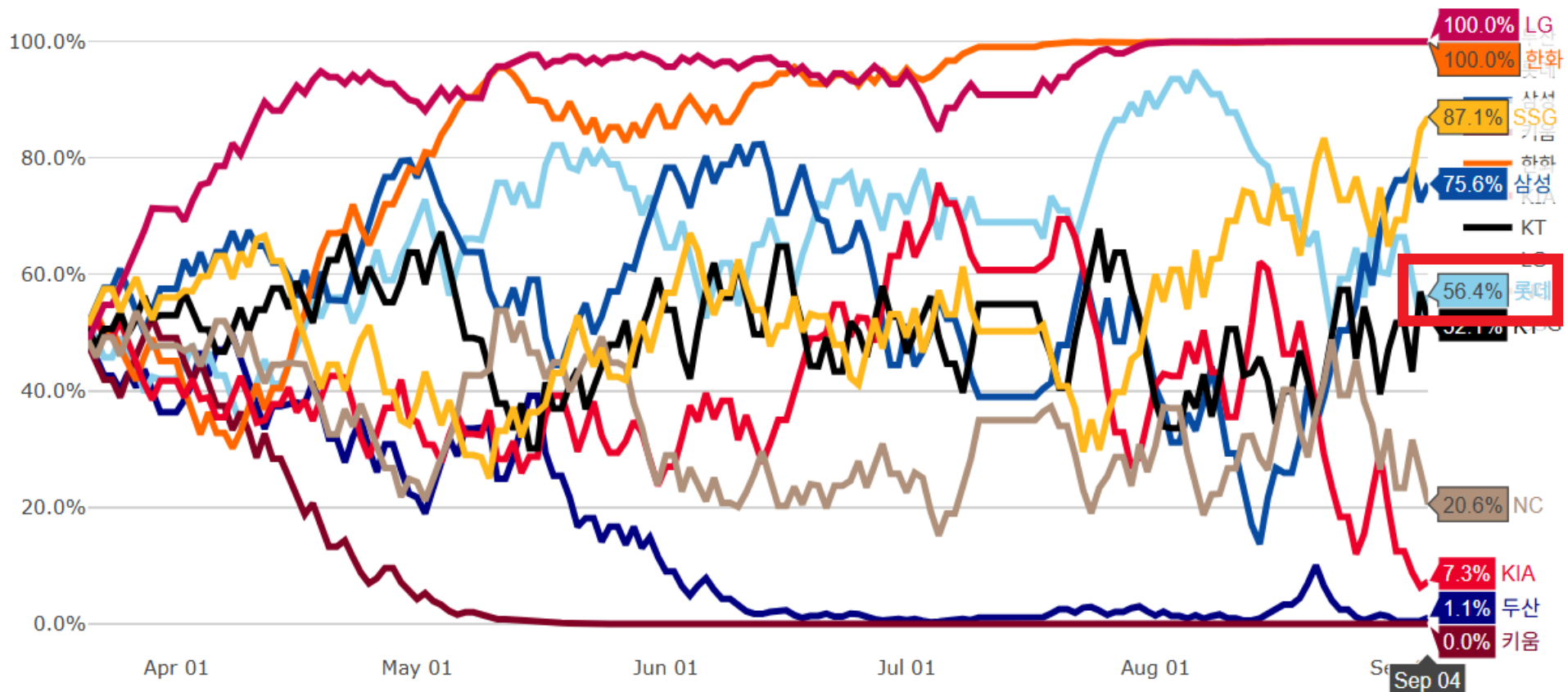
94.9%

1 months

포스트 시즌 진출 확률

56.4%

09/04



포스트 시즌 진출 확률

94.9%

1 months

포스트 시즌 진출 확률

56.4%

Project

연패 탈출 솔루션

Intro

- 필요성
 - 연패 = 전력 관리 실패의 신호
 - 포스트시즌 목표 달성에 치명적
 - 감독/프런트 의사결정 리스크
 - 연패 탈출 솔루션 필요

References

- **KBO 승패 예측 연구**

- 머신러닝·딥러닝으로 경기 결과 및 순위 예측
- 최고 89% 정확도 기록 (딥러닝)
- 규칙 기반 승리 요인 도출 (출루율, 타율, 피안타 등) → 전략 제안

- **해외 MLB 연구 (연승·연패 모멘텀 분석)**

- “모멘텀 효과는 제한적” → 전력 & 상황이 결과 좌우
- 연패 후 승률 = 기본 전력 수준에 수렴
- 에이스 투수, 불펜 상황 등 조건별 확률 차이 존재

- **데이콘(Dacon) 야구 데이터 경진대회**

- OPS 예측 (타자 성적 변동성 예측)
- 외국인 투수 스카우팅 최적화 (KBO·MLB 성적 결합)
- 도메인 지식 + 데이터 과학 융합 사례

Point Of Differentiation

- 분석 대상

- 기존 연구 : **승패 예측** 자체에 집중
- 본 프로젝트 : 연패 구간의 **패턴과 탈출** 조건에 집중

- 데이터 활용 범위

- 기존 연구 : 시즌 전체 성적, **단순 지표**(OPS, OBP, 피안타 등)
- 본 프로젝트 : **팀 수준 고급 지표**(ISO, DER, K/BB, OPS, Baseruns) + streak 구간 데이터 직접 구축

Data

- **타격 지표**

- OPS (출루율+장타율) : 팀 공격력 종합 지표
- ISO (순수 장타력) : 장타 생산 능력 평가
- OBP (출루율) : 득점 기회 창출 능력

- **투수 지표**

- K/BB 비율 : 제구력과 탈삼진 능력 -> 투수 안정성
- DER (수비 효율) : 인플레이 타구 처리 능력

- **팀 지표**

- Baseruns (기대득점 - 실제득점) : 득점 효율성, 운/클러치 상황 반영
- 승률 변화(전/후반기) : 연패 영향 파악

Data Set



스포키

<https://statiz.sporki.com>

[STATIZ \(sporki.com\)](https://statiz.sporki.com)

• 스탯티즈

- 타격/투구
- 팀 지표 활용

팀 기록

Rank	Team	Year	정렬▼ 비율-AVG	G	PA	AB	R	H	2B	3B	HR
1	롯데	21	0.372	1	52	43	15	16	3	0	1
2	삼성	21	0.316	1	45	38	6	12	2	0	0
3	NC	21	0.300	1	38	30	6	9	2	0	1
4	SSG	21	0.270	1	40	37	2	10	0	0	0
5	키움	21	0.219	1	38	32	3	7	1	0	1
5	한화	21	0.219	1	35	32	3	7	1	0	2
5	KIA	21	0.219	1	40	32	4	7	0	0	0
8	LG	21	0.212	1	37	33	3	7	0	0	1
9	두산	21	0.207	1	31	29	3	6	2	1	0
10	KT	21	0.107	1	31	28	0	3	0	0	0

Data Set



스포키

<https://statiz.sporki.com>

[STATIZ \(sporki.com\)](https://statiz.sporki.com)

팀 기록

Rank	Team	Year	정렬▼ 비율-AVG	G	PA	AB	R	H	2B	3B	HR
1	롯데	21	0.372	1	52	43	15	16	3	0	1
2	삼성	21	0.316	1	45	38	6	12	2	0	0
3	NC	21	0.300	1	38	30	6	9	2	0	1
4	SSG	21	0.270	1	40	37	2	10	0	0	0
5	키움	21	0.219	1	38	32	3	7	1	0	1
5	한화	21	0.219	1	35	32	3	7	1	0	2
5	KIA	21	0.219	1	40	32	4	7	0	0	0
8	LG	21	0.212	1	37	33	3	7	0	0	1
9	두산	21	0.207	1	31	29	3	6	2	1	0
10	KT	21	0.107	1	31	28	0	3	0	0	0

• 스탯티즈

- 타격/투구
- 팀 지표 활용

Data Set



스포키

<https://statiz.sporki.com>

[STATIZ \(sporki.com\)](https://statiz.sporki.com)

• 스탯티즈

- 타격/투구
- 팀 지표 활용

										비율			
GO	FO	BB	HP	IB	SO	GDP	SH	SF		AVG▼	OBP	SLG	OPS
12	5	8	1	0	9	1	0	0		0.372	0.481	0.512	0.993
8	7	4	1	1	10	0	1	1		0.316	0.386	0.368	0.754
2	10	6	1	0	6	0	1	0		0.300	0.432	0.467	0.899
3	17	3	0	0	6	0	0	0		0.270	0.325	0.270	0.595
13	5	6	0	0	7	2	0	0		0.219	0.342	0.344	0.686
10	7	3	0	0	7	2	0	0		0.219	0.286	0.438	0.724
12	6	5	3	0	6	1	0	0		0.219	0.375	0.219	0.594
7	11	2	1	0	8	0	1	0		0.212	0.278	0.303	0.581
10	8	2	0	0	4	1	0	0		0.207	0.258	0.345	0.603
15	3	3	0	0	7	2	0	0		0.107	0.194	0.107	0.301

Data Set



스포키

<https://statiz.sporki.com>

[STATIZ \(sporki.com\)](https://statiz.sporki.com)

- 스탯티즈

- 타격/투구
- 팀 지표 활용

										비율		
GO	FO	BB	HP	IB	SO	GDP	SH	SF	AVG▼	OBP	SLG	OPS
12	5	8	1	0	9	1	0	0	0.372	0.481	0.512	0.993
8	7	4	1	1	10	0	1	1	0.316	0.386	0.368	0.754
2	10	6	1	0	6	0	1	0	0.300	0.432	0.467	0.899
3	17	3	0	0	6	0	0	0	0.270	0.325	0.270	0.595
13	5	6	0	0	7	2	0	0	0.219	0.342	0.344	0.686
10	7	3	0	0	7	2	0	0	0.219	0.286	0.438	0.724
12	6	5	3	0	6	1	0	0	0.219	0.375	0.219	0.594
7	11	2	1	0	8	0	1	0	0.212	0.278	0.303	0.581
10	8	2	0	0	4	1	0	0	0.207	0.258	0.345	0.603
15	3	3	0	0	7	2	0	0	0.107	0.194	0.107	0.301

Data Set



KBO

<https://www.koreabaseball.com> > 기록 > 순위 >

타자 | 선수 기록 | 기록실 | 기록/순위 | ...

기록실 · 선수 기록 · 팀 기록 · 투수 VS 타자. 기록용어. 타자 · 투수 · 수비 · 주루. 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994 ...

< 2021.07.06 >



(2021년 07월06일 기준)

정규시즌



• KBO 기록실

• 연패 데이터 활용

순위	팀명	경기	승	패	무	승률	게임차	최근10경기	연속	홈	방문
1	KT	72	44	28	0	0.611	0	8승0무2패	1패	24-0-15	20-0-13
2	LG	75	43	32	0	0.573	2.5	4승0무6패	2승	24-0-16	19-0-16
2	삼성	76	43	32	1	0.573	2.5	5승1무4패	3승	24-0-14	19-1-18
4	SSG	75	40	33	2	0.548	4.5	3승2무5패	1패	20-1-17	20-1-16
5	NC	73	37	34	2	0.521	6.5	5승1무4패	1승	16-1-17	21-1-17
6	키움	78	40	38	0	0.513	7	7승0무3패	2승	22-0-16	18-0-22
7	두산	74	35	39	0	0.473	10	2승0무8패	3패	16-0-20	19-0-19
8	롯데	75	32	42	1	0.432	13	6승0무4패	1패	12-1-22	20-0-20
9	KIA	71	28	43	0	0.394	15.5	4승0무6패	3승	16-0-19	12-0-24
10	한화	75	27	48	0	0.360	18.5	1승0무9패	2패	16-0-23	11-0-25

Data Set



KBO

<https://www.koreabaseball.com> > 기록 > 순위

[타자](#) | [선수 기록](#) | [기록실](#) | [기록/순위](#) | ...

기록실 · 선수 기록 · 팀 기록 · 투수 VS 타자 · 기록용어 · 타자 · 투수 · 수비 · 주루 · 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994 ...

< 2021.07.06 >

(2021년 07월06일 기준)

정규시즌

• KBO 기록실

• 연패 데이터 활용

순위	팀명	경기	승	패	무	승률	게임차	최근10경기	연속	홈	방문
1	KT	72	44	28	0	0.611	0	8승0무2패	1패	24-0-15	20-0-13
2	LG	75	43	32	0	0.573	2.5	4승0무6패	2승	24-0-16	19-0-16
2	삼성	76	43	32	1	0.573	2.5	5승1무4패	3승	24-0-14	19-1-18
4	SSG	75	40	33	2	0.548	4.5	3승2무5패	1패	20-1-17	20-1-16
5	NC	73	37	34	2	0.521	6.5	5승1무4패	1승	16-1-17	21-1-17
6	키움	78	40	38	0	0.513	7	7승0무3패	2승	22-0-16	18-0-22
7	두산	74	35	39	0	0.473	10	2승0무8패	3패	16-0-20	19-0-19
8	롯데	75	32	42	1	0.432	13	6승0무4패	1패	12-1-22	20-0-20
9	KIA	71	28	43	0	0.394	15.5	4승0무6패	3승	16-0-19	12-0-24
10	한화	75	27	48	0	0.360	18.5	1승0무9패	2패	16-0-23	11-0-25

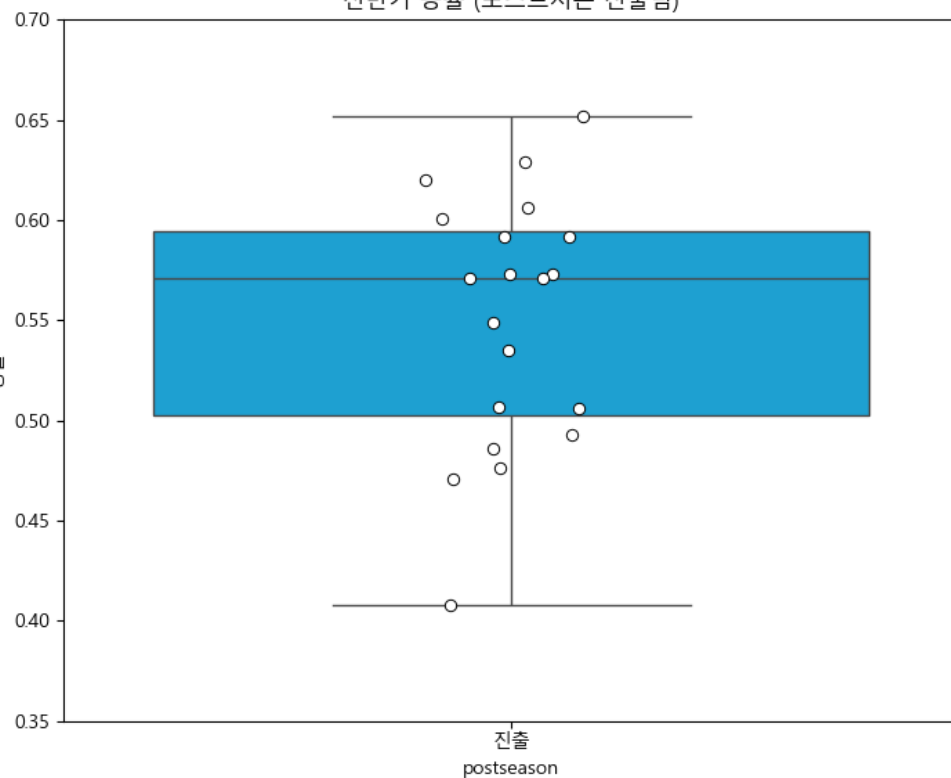
Columns

Rank	팀순위	3B	3루타	IB	고의 4구
Team	팀	HR	홈런	SO	삼진
Year	년도(21)	TB	루타	GDP	병살타
AVG	타율	RBI	타점	SH	희생타
G	경기 수	SB	도루 성공	SF	희생 플라이
PA	타석	CS	도루 실패	Year_	년도(2021)
AB	타수	GO	땅볼 아웃	Date	날짜
R	득점	FO	뜯볼 아웃	OBP	출루율
H	안타	BB	4구	SLG	장타율
2B	2루타	HP	사구	OPS	OBP+SLG

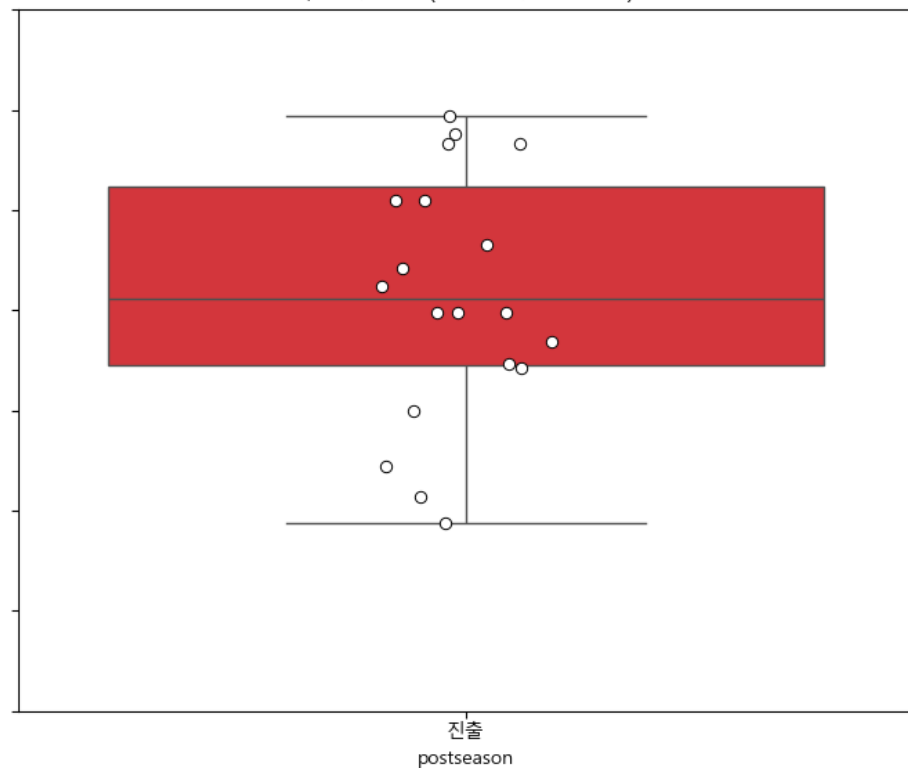
EDA

- 후반기 데이터 선정

전반기 승률 (포스트시즌 진출팀)



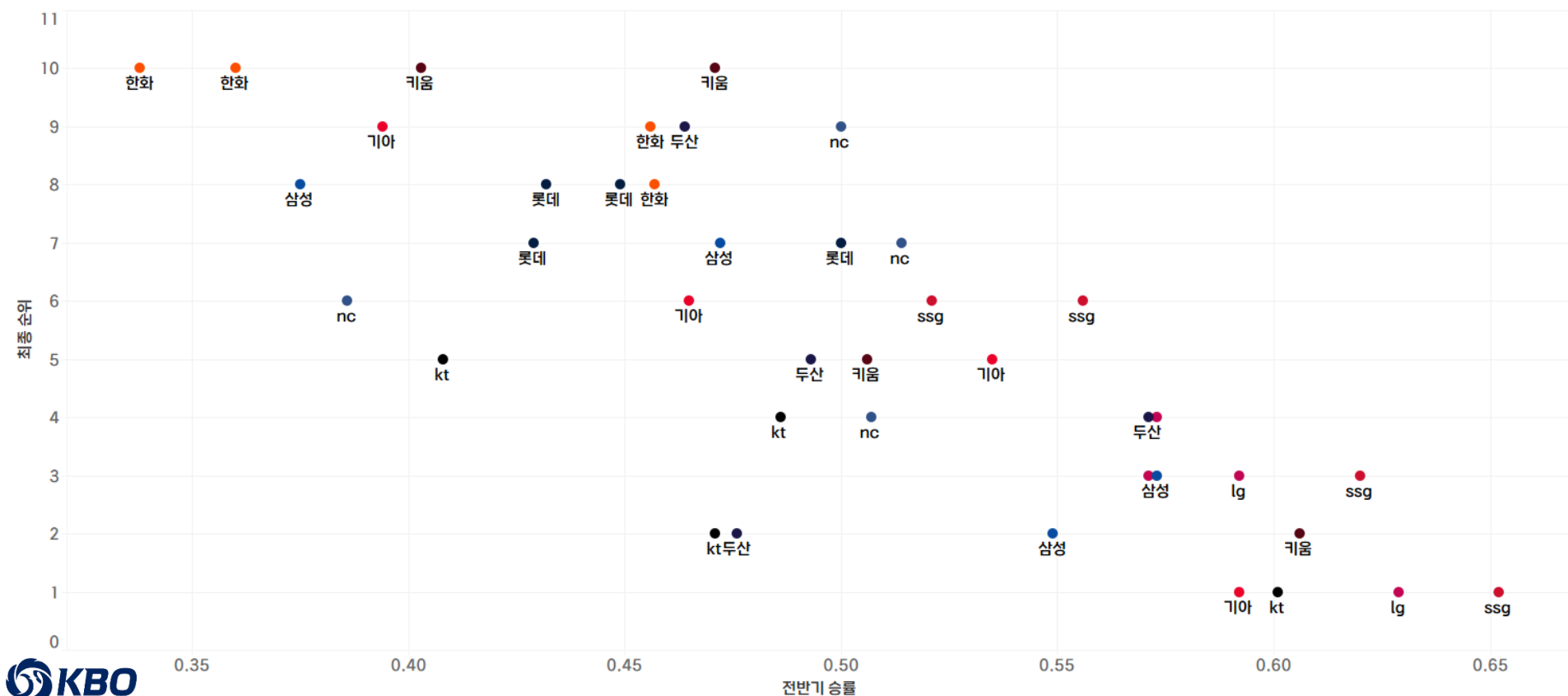
후반기 승률 (포스트시즌 진출팀)



EDA

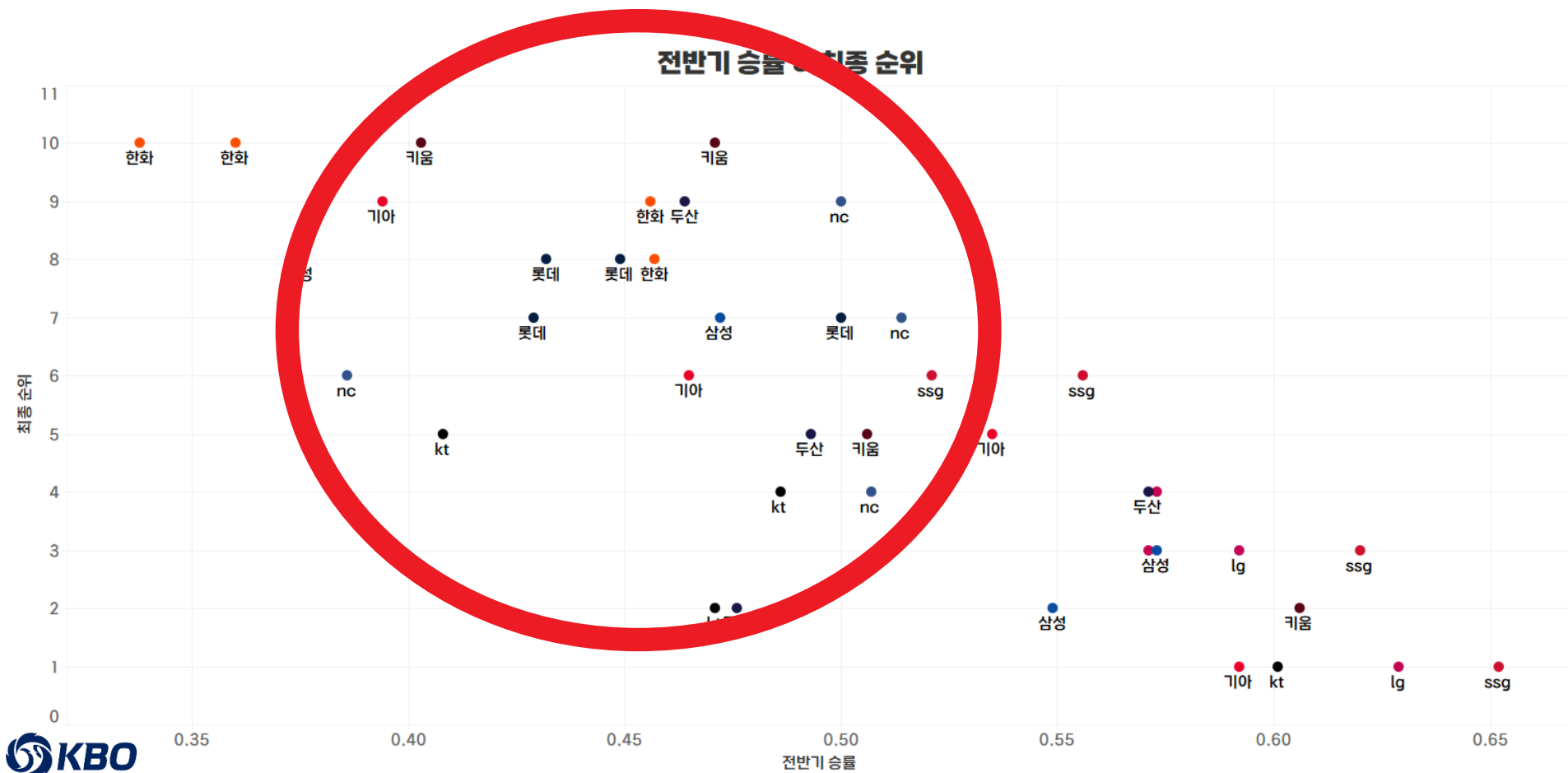
• 후반기 데이터 선정

전반기 승률 & 최종 순위



EDA

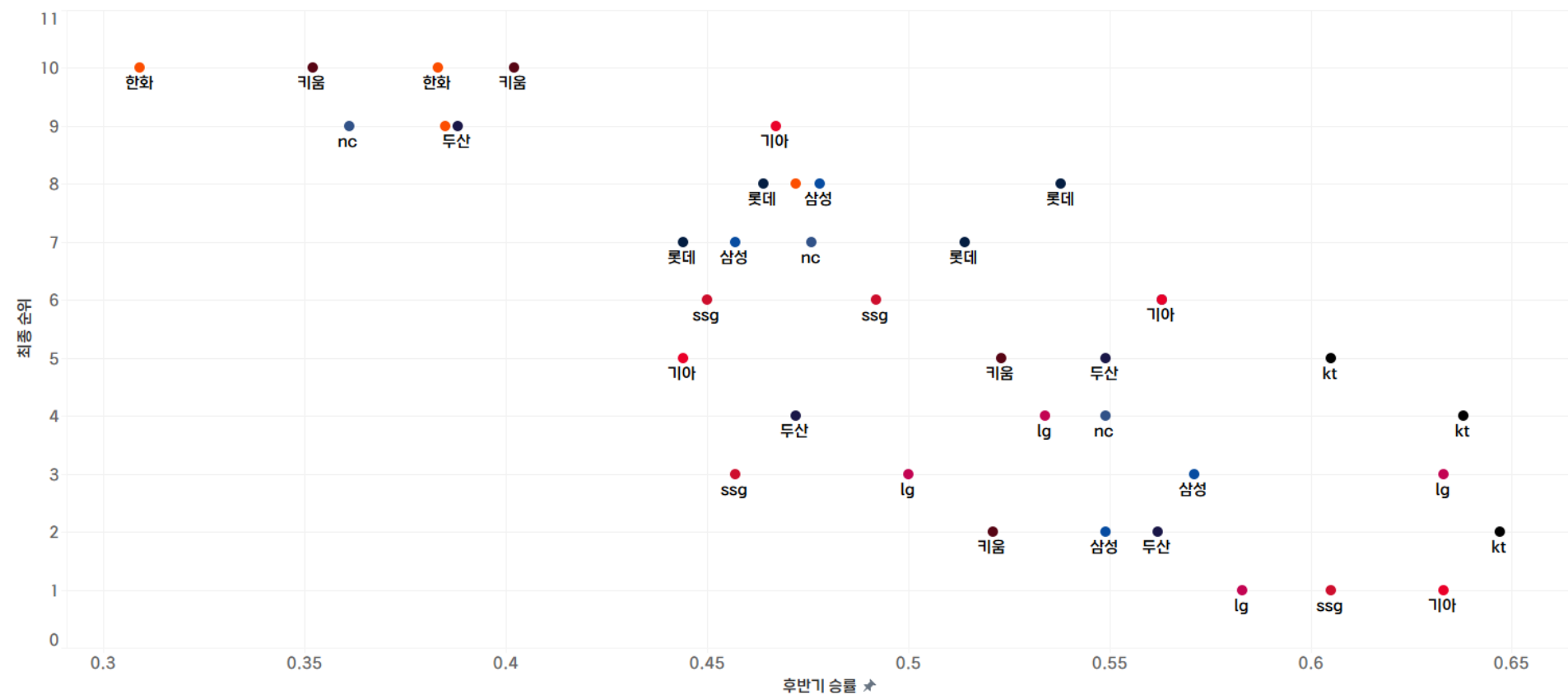
• 후반기 데이터 선정

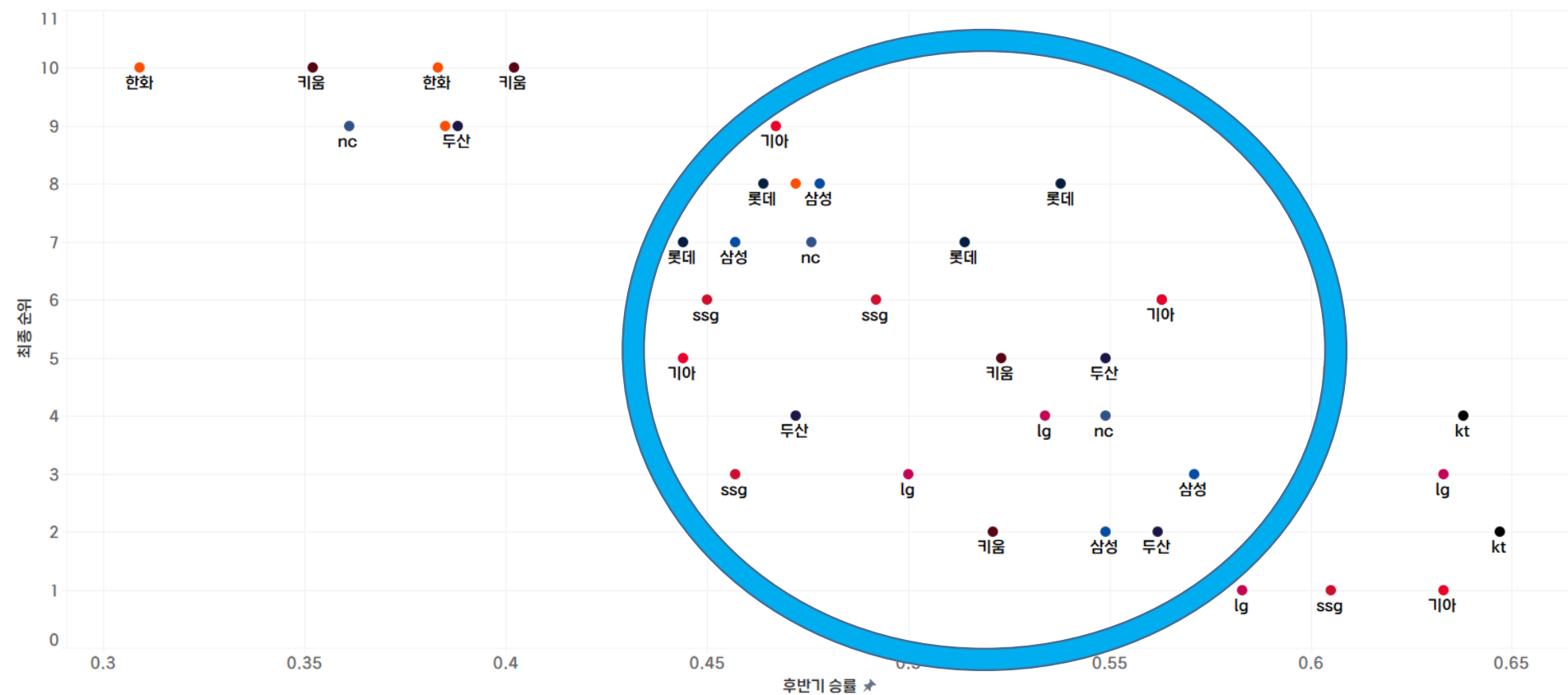


EDA

- 후반기 데이터 선정

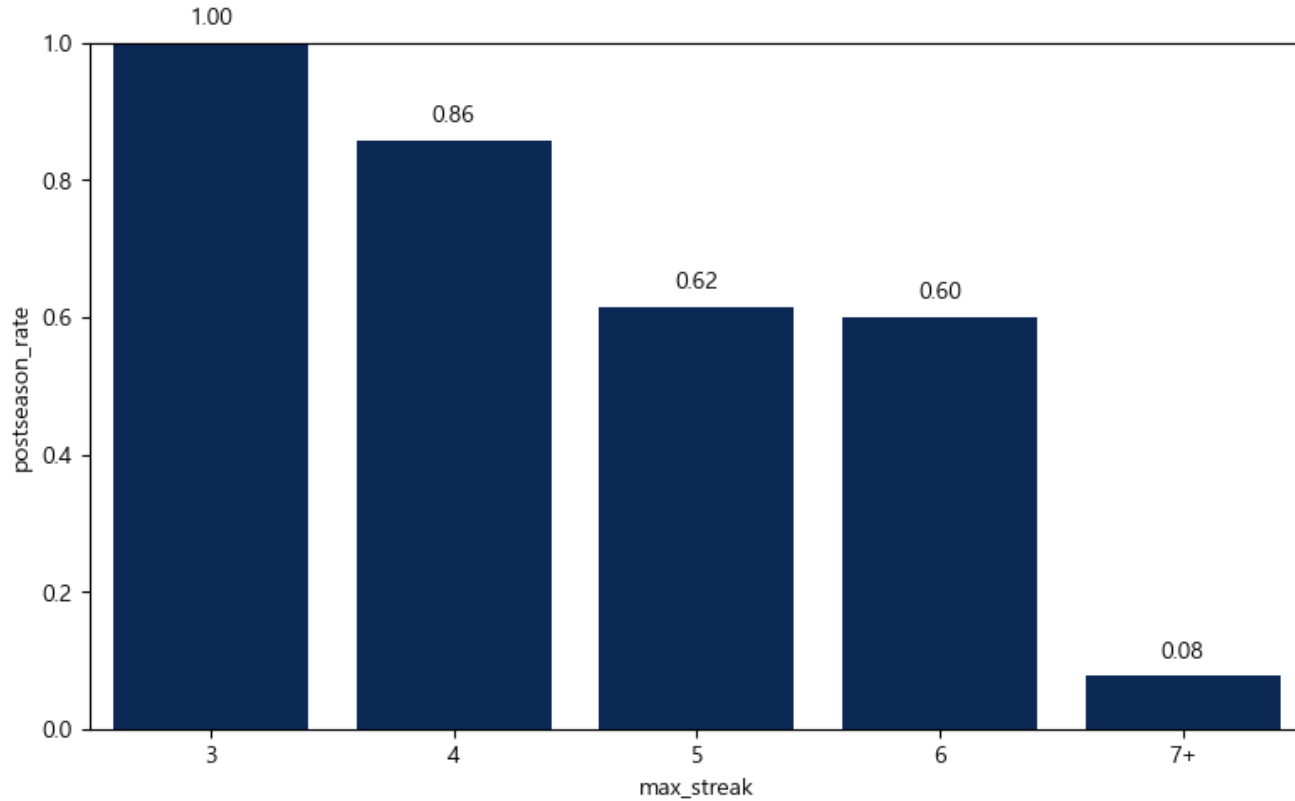
후반기 승률 & 최종 순위





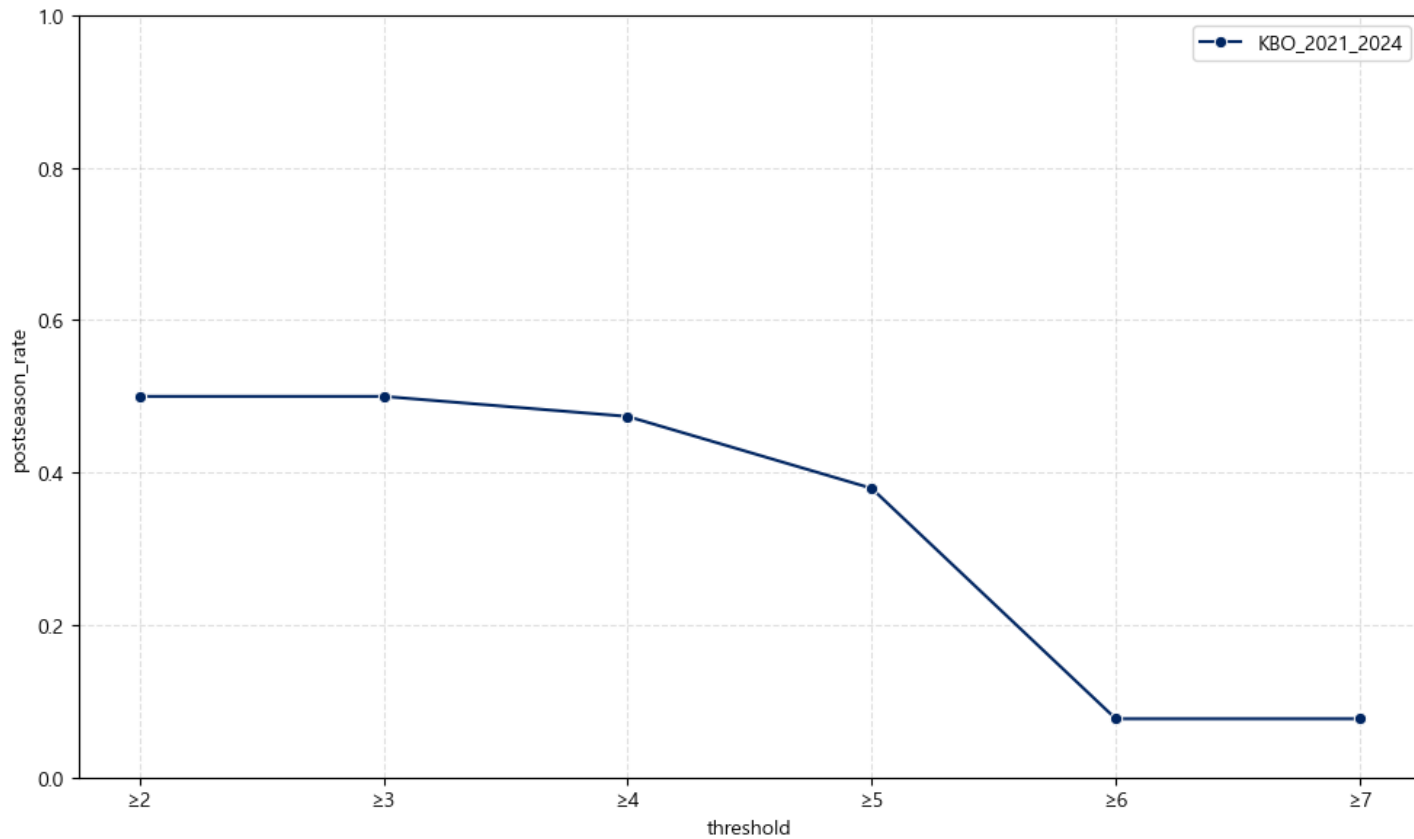
EDA

- 연패 기간 선정



EDA

- 연패 기간 선정



• Pre - Streak - Post

Pre

seg	year	team	...
Pre	2021	두산	...
Pre	2021	두산	...
Pre	2021	두산	...
Pre	2021	두산	...
Pre	2021	두산	...

315 rows × 39 columns

Streak

seg	year	team	...
Streak	2021	두산	...
Streak	2021	두산	...
Streak	2021	두산	...
Streak	2021	두산	...
Streak	2021	두산	...

340 rows × 39 columns

Post

seg	year	team	...
Post	2021	두산	...
Post	2021	두산	...
Post	2021	두산	...
Post	2021	두산	...
post	2021	두산	...

301 rows × 39 columns

• Pre - Streak - Post

구간별 평균 요약:

	R	H	2B	3B	HR	RBI	SB	CS	BB	HP	\
segment											
post	5.296	9.628	1.724	0.176	0.781	4.957	0.591	0.239	3.681	0.568	
pre	5.083	9.486	1.686	0.156	0.746	4.790	0.676	0.244	3.794	0.533	
streak	3.112	7.776	1.262	0.115	0.535	2.909	0.521	0.221	3.150	0.512	
	IB	SO	GDP	SH	SF	AVG	OBP	SLG	OPS		
segment											
post	0.150	7.375	0.731	0.455	0.442	0.270	0.341	0.393	0.734		
pre	0.152	7.324	0.803	0.448	0.337	0.265	0.339	0.383	0.722		
streak	0.091	8.215	0.759	0.359	0.247	0.219	0.290	0.307	0.597		

EDA

- Pre - Streak - Post

결측치 개수:

segment	0
streak_id	0
year	0
team	0
streak_len	0
start_date	0
end_date	0
year_std	0
team_std	0
date_std	0
Rank	0
AVG	0

PA	0
AB	0
R	0
H	0
2B	0
3B	0
HR	0
TB	0
RBI	0
SB	0
CS	0
GO	0
FO	0
BB	0

HP	0
IB	0
SO	0
GDP	0
SH	0
SF	0
AVG_	0
OBP	0
SLG	0
OPS	0
Year_	0
row_order	0
SB_CS	756
dtype:	int64

중복 데이터 개수: 0

• Pre - Streak - Post

수치형 변수 기초 통계량:

	streak_id	year	streak_len	year_std	Rank \
count	956.000000	956.000000	956.000000	956.000000	956.000000
mean	31.731172	2022.627615	5.576360	2022.627615	5.503138
std	18.063643	0.985460	1.963922	0.985460	2.807057
min	1.000000	2021.000000	4.000000	2021.000000	1.000000
25%	16.000000	2022.000000	4.000000	2022.000000	3.000000
50%	31.000000	2023.000000	5.000000	2023.000000	5.000000
75%	47.000000	2023.000000	6.000000	2023.000000	8.000000
max	63.000000	2024.000000	13.000000	2024.000000	10.000000

	AVG	G	PA	AB	R ... \
count	956.000000	956.000000	956.000000	956.000000	956.000000 ...
mean	0.250362	1.018828	39.673640	34.850418	4.448745 ...
std	0.080371	0.135990	7.617435	6.117147	3.493972 ...
min	0.036000	1.000000	19.000000	18.000000	0.000000 ...
25%	0.193000	1.000000	35.000000	32.000000	2.000000 ...
50%	0.250000	1.000000	38.000000	34.000000	4.000000 ...
75%	0.306000	1.000000	42.000000	37.000000	6.000000 ...
max	0.529000	2.000000	90.000000	74.000000	30.000000 ...

• Pre - Streak - Post

	SO	GDP	SH	SF	AVG_	OBP \
count	956.000000	956.000000	956.000000	956.000000	956.000000	956.000000
mean	7.656904	0.764644	0.418410	0.337866	0.250362	0.322279
std	2.950102	0.910696	0.671512	0.591271	0.080371	0.085820
min	0.000000	0.000000	0.000000	0.000000	0.036000	0.037000
25%	5.000000	0.000000	0.000000	0.000000	0.193000	0.264500
50%	8.000000	1.000000	0.000000	0.000000	0.250000	0.324000
75%	10.000000	1.000000	1.000000	1.000000	0.306000	0.381000
max	21.000000	6.000000	4.000000	4.000000	0.529000	0.600000

	SLG	OPS	Year_	row_order
count	956.000000	956.000000	956.000000	956.000000
mean	0.359252	0.681531	2022.627615	90.268828
std	0.138490	0.211057	0.985460	25.707576
min	0.037000	0.074000	2021.000000	38.000000
25%	0.250000	0.523000	2022.000000	69.000000
50%	0.344500	0.668000	2023.000000	89.000000
75%	0.452000	0.830750	2023.000000	109.000000
max	0.857000	1.457000	2024.000000	143.000000

팀별 데이터 개수:

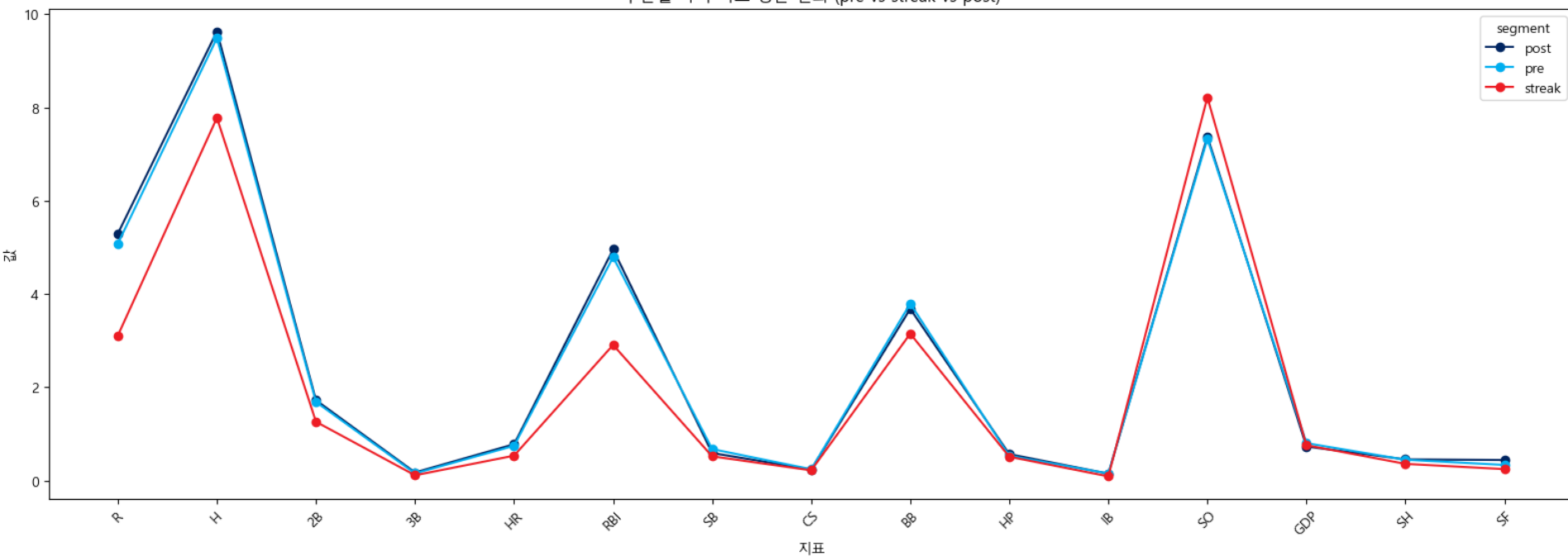
team	
한화	266
키움	231
두산	173
삼성	167
롯데	119

[8 rows x 33 columns]

EDA

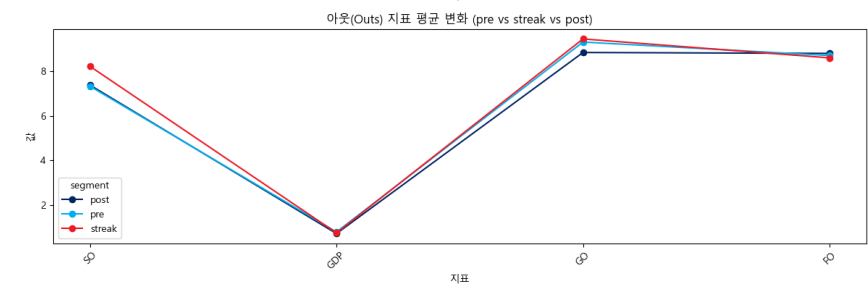
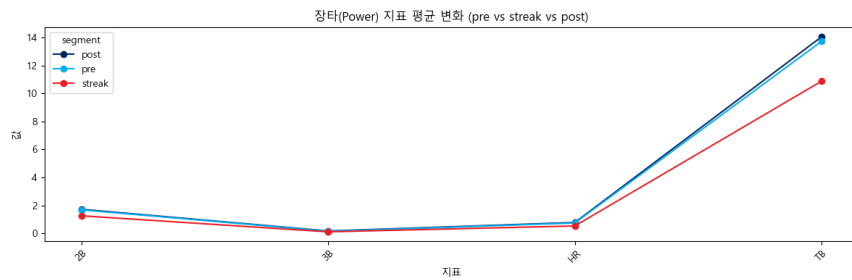
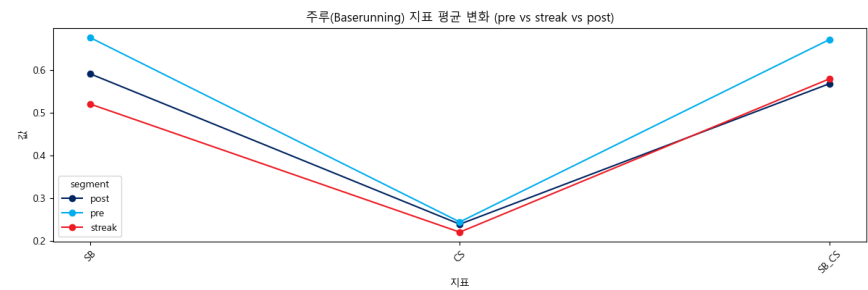
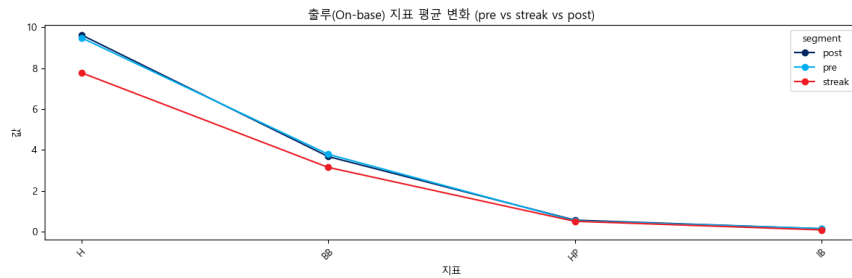
- Pre - Streak - Post (Raw Data)

구간별 타격 지표 평균 변화 (pre vs streak vs post)



EDA

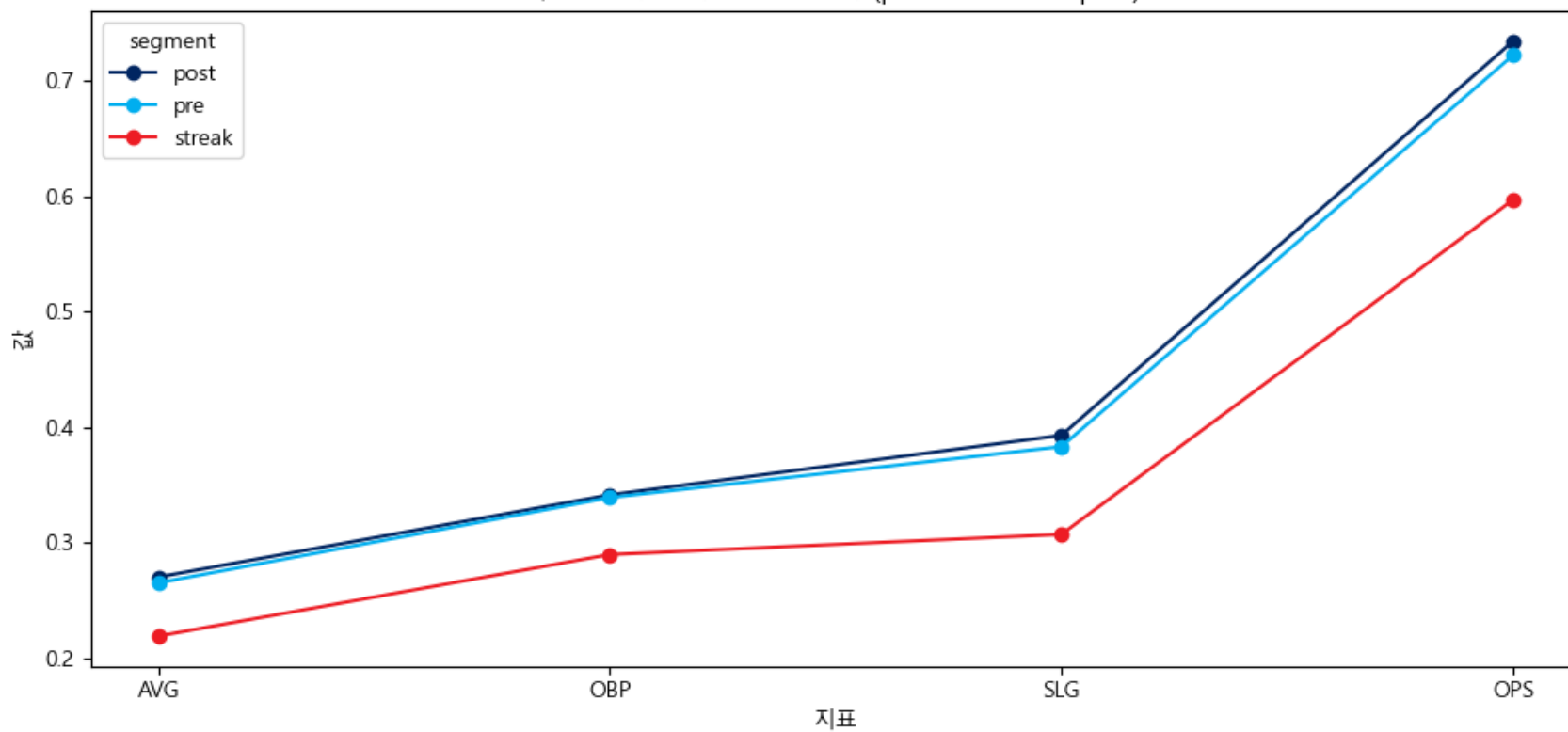
• Pre - Streak - Post (Raw Data)



EDA

- Pre - Streak - Post (Rate Data)

구간별 비율 지표 평균 변화 (pre vs streak vs post)



Modeling

- 고급지표 변환

원시 데이터

H	2B	3B	HR	AB
14	3	2	2	...
8	1	1	4	...
16	5	0	7	...
12	2	4	1	...

ISO	DER	Base_runs	OPS	AB
14	3	2	2	...
8	1	1	4	...
16	5	0	7	...
12	2	4	1	...

Modeling

- 고급지표 변환

고급 지표

H	2B	3B	HR	AB
14	3	2	2	...
8	1	1	4	...
16	5	0	7	...
12	2	4	1	...



ISO	DER	Diff_runs	OPS	AB
14	3	2	2	...
8	1	1	4	...
16	5	0	7	...
12	2	4	1	...

Modeling

- 고급 지표 생성

ISO (순장타율)	$SLG - AVG$
OPS	$OBP + SLG$
OBP (출루율)	$(H + BB + HBP) / (AB + BB + HBP + SF)$
SB_rate (도루 성공률)	$SB / (SB + CS)$
SB+CS (도루 시도 횟수)	$SB + CS$
Diff_runs (기대득점 편차)	Fan_Graphs (MLB 중심 야구 전문 통계 분석 사이트)
DER (수비 효율)	MLB 공식 사이트

Modeling

- 고급 지표 생성

ISO (순장타율)	$SLG - AVG$
OPS	$OBP + SLG$
OBP (출루율)	$(H + BB + HBP) / (AB + BB + HBP + SF)$
SB_rate (도루 성공률)	$SB / (SB + CS)$
SB+CS (도루 시도 횟수)	$SB + CS$
Diff_runs (기대득점 편차)	Fan_Graphs (MLB 중심 야구 전문 통계 분석 사이트)
DER (수비 효율)	MLB 공식 사이트

Modeling

• 해외 References 참고

ISO (순장타율)	$SLG - AVG$
OPS	$OBP + SLG$
OBP (출루율)	$(H + BB + HBP) / (AB + BB + HBP + SF)$
SB_rate (도루 성공률)	$SB / (SB + CS)$
SB+CS (도루 시도 횟수)	$SB + CS$
Diff_runs (기대득점 편차)	Fan_Graphs (MLB 중심 야구 전문 통계 분석 사이트)
DER (수비 효율)	MLB 공식 사이트

Modeling

- 변수 선택 근거

변수	Shapiro-Wilk (pre)	Shapiro-Wilk (streak)	Shapiro-Wilk (post)	Levene (등분산성)
ISO	0.000	0.000	0.000	0.001
DER	0.000	0.000	0.000	0.147
SB_CS	0.000	0.000	0.000	0.467
SB_rate	0.000	0.000	0.000	0.126
diff_runs	0.001	0.001	0.000	0.016
OBP	0.730	0.160	0.602	0.042
OPS	0.089	0.005	0.190	0.001

Modeling

• 변수 선택

변수	정규성 (Shapiro-Wilk) p값	등분산성 (Levene) p값	선택된 검정 방법	검정 통계량	p값	해석
ISO	pre=0.000 / streak=0.000 / post=0.000	0.001	Kruskal-Wallis	30.709	0.000	세 집단 간 분포 차이 있음
DER	pre=0.000 / streak=0.000 / post=0.000	0.147	Kruskal-Wallis	65.322	0.000	세 집단 간 분포 차이 있음
SB_CS	pre=0.000 / streak=0.000 / post=0.000	0.467	Kruskal-Wallis	6.655	0.036	세 집단 간 분포 차이 있음 (경계적)
SB_rate	pre=0.000 / streak=0.000 / post=0.000	0.126	Kruskal-Wallis	4.005	0.135	유의한 차이 없음
OPS	pre=0.089 / streak=0.005 / post=0.190	0.001	Kruskal-Wallis	86.805	0.000	세 집단 간 분포 차이 있음

Modeling

- 분류 문제

ISO	DER	KBB	OBP	OPS	Diff_runs	target_streak
0.179	0.567	0.25	0.511	1.126	-0.28	0
0.091	0.931	1.166	0.25	0.508	-7.6	0
0	1	1.2	0.37	0.597	0.093	0
0.17	1	4.333	0.378	0.866	-3.22	1
0.128	1	3.5	0.235	0.525	0.24	1

Modeling

- 분류 문제

ISO	DER	KBB	OBP	OPS	Diff_runs	target_streak
0.179	0.567	0.25	0.511	1.126	-0.28	0
0.091	0.931	1.166	0.25	0.508	-7.6	0
0	1	1.2	0.37	0.597	0.093	0
0.17	1	4.333	0.378	0.866	-3.22	1
0.128	1	3.5	0.235	0.525	0.24	1

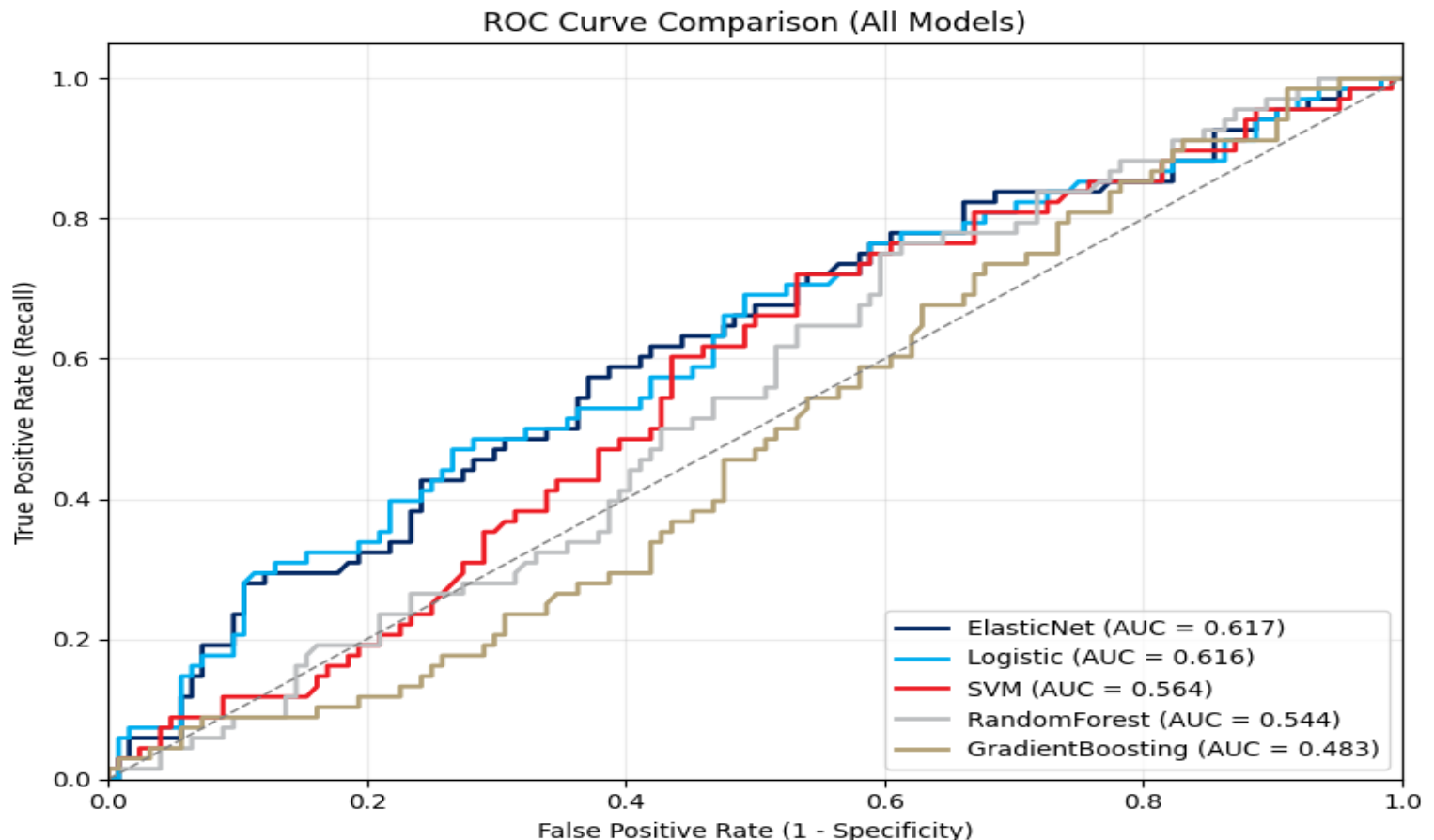
Modeling

- 분류 문제

ISO	DER	KBB	OBP	OPS	Diff_runs	target_streak
0.179	0.567	0.25	0.511	1.126	-0.28	0
0.091	0.931	1.166	0.25	0.508	-7.6	0
0	1	1.2	0.37	0.597	0.093	0
0.17	1	4.333	0.378	0.866	-3.22	1
0.128	1	3.5	0.235	0.525	0.24	1

Model Evaluation

- ROC 곡선



Model Evaluation

• 튜닝 결과

항목	값	설명
Penalty	L1	L1 규제(라쏘). 일부 계수를 0으로 만들어 변수 선택 효과 있음
Solver	liblinear	작은 데이터셋에서 L1/L2 지원하는 최적화 알고리즘
C	2	값이 클수록 규제 약함. → 중간 수준 규제 선택
Best CV ROC-AUC	0.7118	랜덤(0.5)보다 유의미하게 높음, 약 71% 순위 구분 정확도

Model Evaluation

- Threshold

항목	threshold	설명
Youden's J (TPR - FPR 최대화)	0.6092	민감도와 특이도의 균형
F1-score 최대화	0.3819	Precision과 Recall의 균형

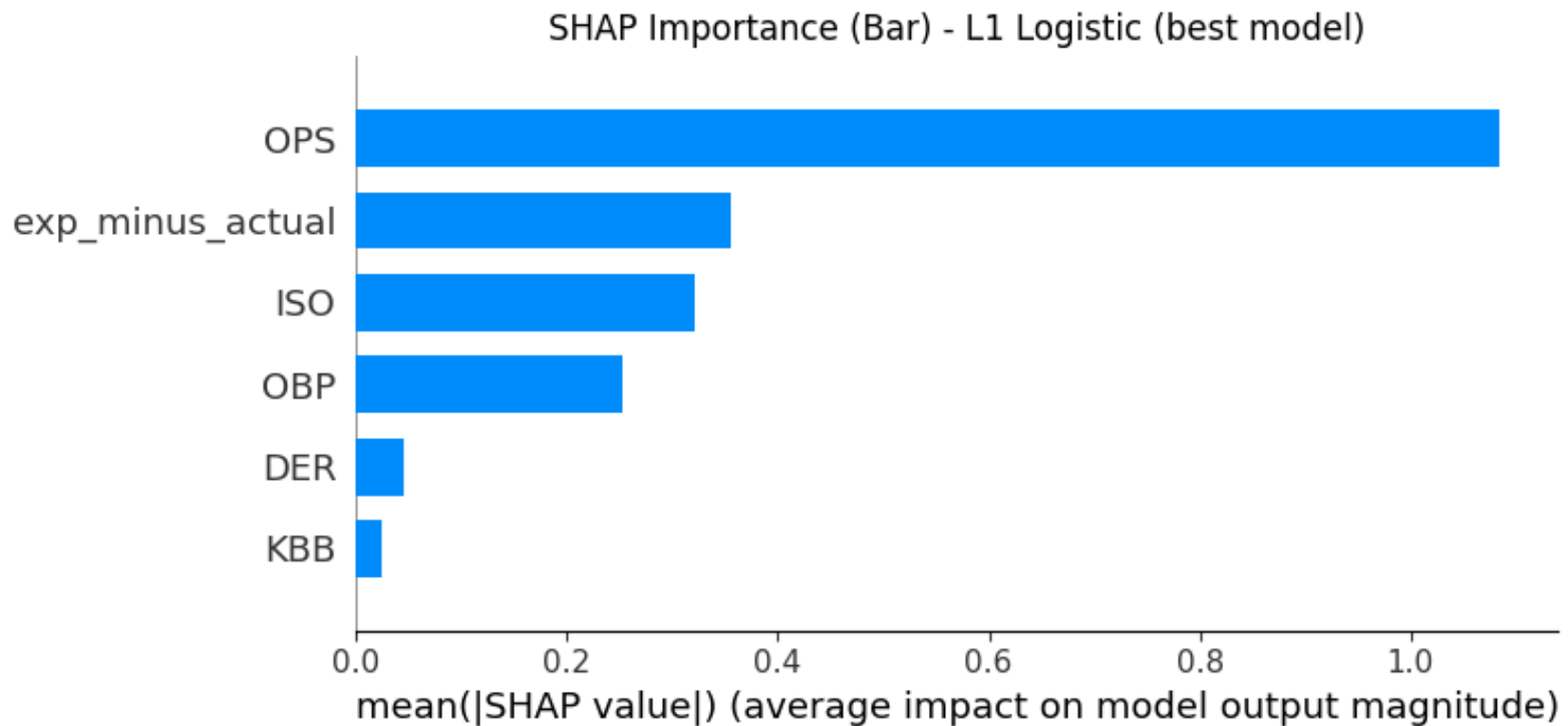
Model Evaluation

- Threshold

항목	threshold	설명
Youden's J (TPR - FPR 최대화)	0.6092	민감도와 특이도의 균형
F1-score 최대화	0.3819	Precision과 Recall의 균형

Insight

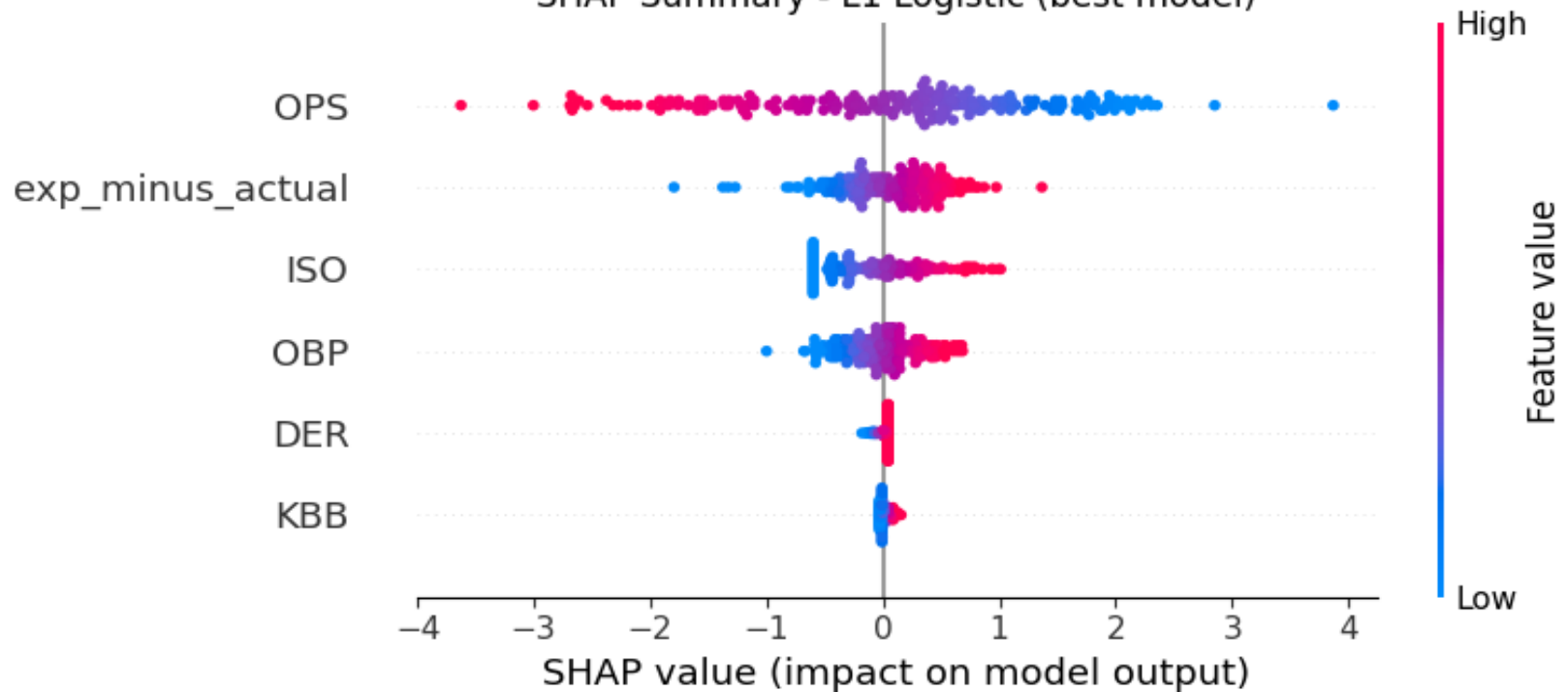
- 연패 핵심 요인, 탈출 조건 요약



Insight

- 연패 핵심 요인, 탈출 조건 요약

SHAP Summary - L1 Logistic (best model)



Result

• Test

```
import joblib
import pandas as pd
# 08-14(롯데-6연패 _12연패 중심)
# 1) 저장된 모델 불러오기
model = joblib.load("logistic_streak_model.pkl")

# 2) 새로운 경기 데이터 입력
new_game = pd.DataFrame([{"ISO": 0.042,
                           "DER": 0.875,
                           "KBB": 4.00,
                           "OBP": 0.333,
                           "OPS": 0.708,
                           "exp_minus_actual": -0.666
                           }])

# 3) 확률 예측
prob = model.predict_proba(new_game)[:, 1][0]

# 4) threshold 적용
pred = (prob >= 0.3).astype(int)

print("연패 확률:", round(prob, 3))
print("예측 결과:", "연패" if pred==1 else "비연패")
```

연패 확률: 0.321
예측 결과: 연패

```
import joblib
import pandas as pd

# 08-24 (롯데 연패 탈출 시점)
# 1) 저장된 모델 불러오기
model = joblib.load("logistic_streak_model.pkl")

# 2) 새로운 경기 데이터 입력
new_game = pd.DataFrame([{"ISO": 0.299,
                           "DER": 0.833,
                           "KBB": 1.375,
                           "OBP": 0.472,
                           "OPS": 1.123,
                           "exp_minus_actual": -4.059
                           }])

# 3) 확률 예측
prob = model.predict_proba(new_game)[:, 1][0]

# 4) threshold 적용
pred = (prob >= 0.3).astype(int)

print("연패 확률:", round(prob, 3))
print("예측 결과:", "연패" if pred==1 else "비연패")
```

연패 확률: 0.082
예측 결과: 비연패

Result

• Test

```
import joblib
import pandas as pd
# 08-14(롯데-6연패 _12연패 중심)
# 1) 저장된 모델 불러오기
model = joblib.load("logistic_streak_model.pkl")

# 2) 새로운 경기 데이터 입력
new_game = pd.DataFrame([{"ISO": 0.042,
                           "DER": 0.875,
                           "KBB": 4.00,
                           "OBP": 0.333,
                           "OPS": 0.708,
                           "exp_minus_actual": -0.666
                           }])

# 3) 확률 예측
prob = model.predict_proba(new_game)[0, 1][0]

# 4) threshold 적용
pred = (prob >= 0.3).astype(int)

print("연패 확률:", round(prob, 3))
print("예측 결과:", "연패" if pred==1 else "비연패")
```

연패 확률: 0.321
예측 결과: 연패

```
import joblib
import pandas as pd

# 08-24 (롯데 연패 탈출 시점)
# 1) 저장된 모델 불러오기
model = joblib.load("logistic_streak_model.pkl")

# 2) 새로운 경기 데이터 입력
new_game = pd.DataFrame([{"ISO": 0.299,
                           "DER": 0.833,
                           "KBB": 1.375,
                           "OBP": 0.472,
                           "OPS": 1.123,
                           "exp_minus_actual": -4.059
                           }])

# 3) 확률 예측
prob = model.predict_proba(new_game)[0, 1][0]

# 4) threshold 적용
pred = (prob >= 0.3).astype(int)

print("연패 확률:", round(prob, 3))
print("예측 결과:", "연패" if pred==1 else "비연패")
```

연패 확률: 0.082
예측 결과: 비연패

Limitations

- 한계점
 - 데이터 범위와 표본 크기
 - 변수(Feature) 한정성
 - 모델링 성능

What's Next?

- Develop
 - 데이터 확장
 - 고급 변수 추가
 - 모델링 고도화

감사합니다.